

SUBJECTIVE ASSESSMENT OF THE IMPACT OF TRANSMISSION ERRORS IN 3DTV COMPARED TO HDTV

Jesús Gutiérrez,[†] Pablo Pérez,[§] Fernando Jaureguizar,[†] Julián Cabrera,[†] and Narciso García[†]

[†]Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

[§]Alcatel-Lucent, María Tubau 9, 28050 Madrid, Spain.

ABSTRACT

Recently, broadcasted 3D video content has reached households with the first generation of 3DTV. However, few studies have been done to analyze the Quality of Experience (QoE) perceived by the end-users in this scenario. This paper studies the impact of transmission errors in 3DTV, considering that the video is delivered in side-by-side format over a conventional packet-based network. For this purpose, a novel evaluation methodology based on standard single stimulus methods and with the aim of keeping as close as possible the home environment viewing conditions has been proposed. The effects of packet losses in monoscopic and stereoscopic videos are compared from the results of subjective assessment tests. Other aspects were also measured concerning 3D content as naturalness, sense of presence and visual fatigue. The results show that although the final perceived QoE is acceptable, some errors cause important binocular rivalry, and therefore, substantial visual discomfort.

Index Terms— Quality of experience, 3DTV, subjective evaluation, transmission errors, visual discomfort

1. INTRODUCTION

Since the recent arrival of 3D video services to the general public, the popularity of this type of audiovisual content has greatly increased. This fact, in addition to an important technology development, has allowed 3D video to reach also the home environment. Nowadays, people can buy a 3D television set to watch stereoscopic content in their households. However, the available content is usually restricted to that locally stored in media like Blu-ray. Therefore, the next challenge was to deliver 3D sequences to end-users by broadcasting, which has been accomplished recently with the appearance of the first channels offering 3D programs.

Nevertheless, delivering 3D video is still a very active field of research, since better QoE perceived by the end-users should be achieved. This implies getting high video quality, good depth perception and naturalness, without causing visual discomfort to the observers. With this aim, several studies have been done concerning the processes of capture, encoding, transmission and display of 3D video contents [1]. The existing and new proposals for improving 3DTV performance should be evaluated with QoE assessment techniques.

While several quality evaluation methods for monoscopic video have been proposed in the literature [2], the first steps are being done in QoE assessment for 3D content. In the case of subjective assessment techniques, new aspects should be taken into account to evaluate properly all the factors involved in 3D video quality [3]. Related

to objective quality assessment, researchers are developing the first proposals, generally based on 2D metrics [4].

In this paper a study of the impact of transmission errors in 3DTV is presented. In contrast to the numerous studies of this type of distortions made for conventional video [5][6], only a few publications exist for stereoscopic sequences [7][8]. Moreover, little research has been done analyzing the performance of the first approaches for delivering 3DTV that are being recently used [9]. Therefore, an extensive study of video and audio distortions considering a broadcasting scenario based on side-by-side 3D video was carried out. In addition, monoscopic sequences were used to compare the effects of transmission errors. Furthermore, other factors related to the 3D viewing experience were evaluated, like naturalness, sense of presence and visual discomfort.

This paper is organized as follows. In Section 2, the broadcasting architecture and the transmission errors considered in the study are described. The experimental setup of the subjective tests is detailed in Section 3. The obtained results are presented in Section 4 and the final conclusions in Section 5.

2. 3DTV BROADCASTING ARCHITECTURE

The first approaches for 3D video broadcasting are focused on profiting from the existing infrastructure used for delivering monoscopic video [9]. Specifically two main techniques are being used: frame sequential simulcast and spatial multiplex of left and right views. The first method is based on encoding left and right views separately and transmitting them consecutively, which implies the transmission of the double of data compared to the case of monoscopic video. This fact usually causes a major drawback for broadcasters, therefore spatial multiplexing of both left and right views into one conventional frame is used. This frame is encoded as a single image and transmitted, without increasing the amount of data in comparison with monoscopic video broadcasting. However, this technique has the disadvantage of losing image resolution due to the merging of two full images into one single frame. Various methods for mixing the stereoscopic views into one frame have been proposed, but the most common is Side-by-Side (SbS), in which the width of left and right views is reduced to the half and both images are placed one next to the other.

In this paper, a typical broadcasting architecture is considered in which a first step for multiplexing left and right frames has been added, as shown in Figure 1. The specific technique considered is SbS multiplexing and the generated frames are encoded using



Fig. 1. 3DTV broadcasting chain.

H.264/AVC as in monoscopic videos. Then, the compressed stream is transmitted to the end-user side, where it is decoded. Finally, the left and right views are separated and displayed in a 3D screen.

The transmission of the video over a lossy network can cause several distortions affecting to video and audio, entailing degradations of the quality perceived by the end-user. Specifically, the errors described in Table 1 were considered in this paper.

Error type	Description
R	Bitrate drops.
F	Framerate drops.
E	Video losses producing macroblocking. The losses could affect different fractions of the frames and various lengths were considered.
V	Video freeze of different duration.
A	Audio losses of different duration.
AV	Video freeze combined with audio loss.

Table 1. Considered transmission errors.

3. EXPERIMENTAL SETUP

3.1. Environment and Equipment

The subjective tests were carried out at the Universidad Politécnica de Madrid, in a lab especially fitted-out to satisfy the recommendations given in ITU-R BT.500-11 [10] for assessments in home environments. The display used for presenting the sequences was a 42" Panasonic stereoscopic television, with resolution of 1920x1080 and aspect ratio 16:9. For visualizing 3D sequences, the observers wore the active shutter glasses associated to this TV set. The subjects were positioned at a viewing distance of 3 times the height of the TV. The videos used in the assessment sessions were played using a PC and the Nvidia 3D Vision Video Player for displaying stereoscopic and monoscopic video.

3.2. Observers

In the subjective experiment participated 19 observers (2 female, 17 male), all of them having normal visual and stereoscopic acuity and color vision. The ages of the subjects were ranged between 24 and 48 years old, with an average age of 31. No observers were rejected after the screening of the subjective results, according to the recommendation ITU-R BT.500-11 [10] to discard subjects who could have voted randomly.

3.3. Test material

Two different video sources of 5 minutes of duration were used in the subjective tests, whose main characteristics are shown in Table 2.

To introduce transmission errors and evaluate them, these sequences were firstly encoded in H.264/AVC. A GOP length of 24

Source	Format	Content
1	1920x1080p 23.97fps	Movie. Some slow segments with dialogs. Some others with fast camera movement.
2	720x576p 25fps	Documentary. Slow action. Some segments with camera panning. Only music as soundtrack.

Table 2. Test sequences

frames was used following the structure IBBBP, and setting all I-frames as Instant Decoder Refresh (IDR) pictures. The bitrate used for encoding the videos were 8Mbps for video source 1, and 4Mbps for the second one. After that, they were divided into segments of around 6 seconds, in order to insert transmission errors randomly in alternate segments. A number was also printed in the right-bottom corner of the image in each undistorted segment, being a reference for the users to do the evaluation, as it will be explained in Subsection 3.4. All the segments were stitched together again, obtaining continuous sequences.

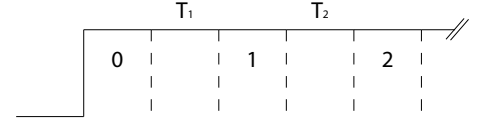


Fig. 2. Diagram of the final structure of the test sequence.

The final structure of each test sequence is shown in Figure 2. The first segment of the sequences, lacking any errors, has a 0 label triggering the start of the viewing time and providing a reference of coding quality of the video. In the following segments, during each segment with a printed number, the previous one (T_i) is evaluated. This preparation of the content permit a nearly continuous evaluation without losing the continuity of the video.

3.4. Procedure

The evaluation tests were carried out organizing different assessment sessions of less than 30 minutes. A maximum of 4 subjects took part in each assessment session. Each session started with an explanation of the tests focused on clarifying the objective of the experiments and familiarizing the subjects with the evaluation methodology. Therefore, a sample sequence, different to those used for evaluation, was presented as an example containing the possible transmission errors. After the explanation, the monoscopic version of video source 1 was evaluated firstly, followed by the rating of its stereoscopic version. Finally, the monoscopic and stereoscopic versions of video source 2 were rated.

To evaluate the impact of transmission errors on the quality perceived by the end-user of a broadcasting system, the assessment method should keep as close as possible the typical viewing conditions in real home environments. Furthermore, the effects of this type of distortions on video quality is highly dependent of time aspects, like its duration or the moment when it happens, so it is necessary to evaluate the quality of the sequence in a continuous way. Due to these reasons, the assessment method used in the tests was based on the Single Stimulus Continuous Quality Evaluation [10].

The observers had to evaluate the sequences (monoscopic and stereoscopic) generated from the source videos as explained in Section 3.3, using the five-grade impairment scale (5 imperceptible, 4 perceptible but not annoying, 3 slightly annoying, 2 annoying and 1 very annoying) defined in [10]. According to the numbers that appear in the sequence, the subjects had to fill a questionnaire which shows a table with the rating scale and 25 squares (one for each distorted segment). When a number i (except 0) appears in the screen, they have to give a score for the impact of the distortion in the previous segment T_i (as shown in Figure 2), writing a cross in the corresponding cell. As the segments with printed numbers have no errors, the observers can avert their gaze to the questionnaire if needed. The proposed assessment methodology is similar to the Absolute Category Rating described by VQEG in [11], but instead of inserting grey segments to make the evaluation, segments with no degradation are introduced to maintain the continuity of the sequence and get closer

to real home viewing conditions. As each segment is different to the others, no explicit reference is used in the evaluation.

After the evaluation of the segments of each stereoscopic video, the observers were asked to indicate their preference between this one and the corresponding monoscopic version. In order to measure the performance of 3D content, the subjects also evaluated the naturalness, the sense of presence and the degree of visual fatigue. The five-grade quality scale (5 excellent, 4 good, 3 fair, 2 poor, 1 bad) was used to rate the naturalness and sense of presence, while the fatigue was measured considering the five-grade scale defined in [12] (ranged from: 5 "My eyes are not tired" to 1 "My eyes are tired"). In the questionnaire, there was also space available for the observers to freely describe other types of discomfort felt during the tests, like headache or dizziness.

4. RESULTS

4.1. Impact of transmission errors

The scores provided by the observers during the test sessions were collected and grouped by types of error, according to the description presented in Table 1. Then, the MOS values for each type of error were calculated to study their impact on perceived quality. A 95% confidence interval for these values was also computed as described in [10]. A comparison of the global results obtained for monoscopic and stereoscopic visualizations of both source videos is shown in Figure 3. Comparing the effects of the different transmission errors analyzed, it can be seen that a similar trend is obtained for conventional video and 3DTV. As it can be expected, the most annoying distortion is the loss of audio together with video freeze (AV), as the observers miss all the multimedia information. Analyzing the effects of the loss of video and audio packets, it can be seen that the audio losses (A) are more annoying than video losses (E), when at least some video content is correctly displayed. However, when the video losses cause video freeze (V) the subjective quality is worse in comparison to that perceived during audio losses (A).

Bitrate and framerate drops are typical effects that take place when the performance of the network decrease. The results show that a bitrate drop (R) is preferred to a framerate drop (F). The main cause is that the latter breaks the natural flow of the video causing usually more annoyance than the video quality reduction produced by the former.

Comparing the effects of transmission errors for monoscopic and stereoscopic content, very similar results are shown in Figure 3. Therefore, a more extensive statistical analysis has been done to find out whether there is statistical significance between the results for conventional video and 3DTV. Firstly, the rating results provided by the observers for each error were grouped in one sample for monoscopic video and another for 3D video (two samples for each distortion). Then, the Kolmogorov-Smirnov test [13] was applied to find out that the samples were not normally distributed at the 5% significance level. Thus, non-parametric methods for comparing each two samples were considered. Specifically, the Wilcoxon Signed-Rank test [13] was used.

One of the main aspects to analyze is the effect of video losses (E) in frame-compatible 3DTV, where the broadcasting system for conventional video is used. This implies that the frame containing the stereo views is encoded using H.264/AVC, and when a video packet is lost, the error concealment technique of the decoder acts without taking into account that it is stereoscopic video. This fact may cause that the blocking effects caused by video losses appear in different regions of each stereo view. Therefore, very different content could appear in corresponding regions of each view. Then,

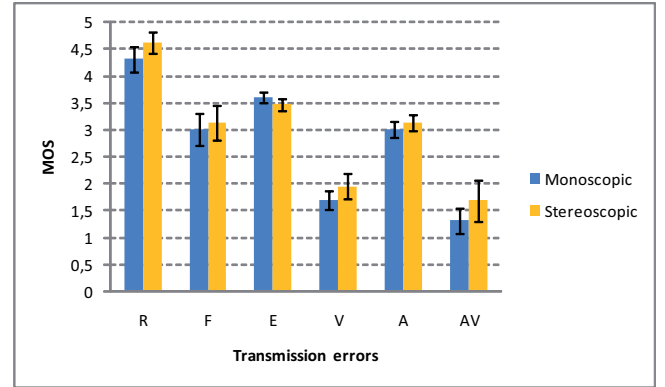


Fig. 3. Comparison of the impact of transmission errors.

when the 3D video is displayed, those content differences could difficult the fusion carried out by the Human Visual System (HVS) of both stereo views, causing visual discomfort to the observers. This is reflected in the better MOS results obtained for monoscopic video affected by video losses (E) in comparison with 3D video, which have statistical significance at the 1% level (Wilcoxon test result: $z=2.927$). Similar conclusions were reported in [7] in a simulcast broadcasting scenario.

In contrast, better results were obtained for 3D video for the rest of the transmission errors. Results from the Wilcoxon tests showed statistical significance for the comparisons related to bitrate drops ($T=63$, 5% level), audio losses ($z=2.556$, 2% level) and video freeze ($z=2.691$, 1% level). This could be attributed to the added value provided by the factors related to the depth perception of 3DTV. However, no statistical significance was found in the comparison of the results for framerate drops (F) and video freeze combined with audio loss (AV). In the first case, this is probably due to the annoyance of the alteration of the natural video pace, while in the latter, the loss of all the information of the audiovisual content could be the cause.

4.2. 3D performance

The performance of the 3D video system used in the experiments is analyzed from the results of the evaluation of naturalness, sense of presence and discomfort. In Figure 4 the obtained MOS values and the 95% confidence interval for both source videos are presented. The results show that these factors are highly dependent on the video content and the creation and post-production of the video. As it can be seen, the first video source satisfies better the observers expectations for 3D video, obtaining greater values of naturalness and sense of presence.

A critical factor related to the experience of viewing 3D video which is being extensively studied is the visual discomfort felt by the observers [7][12]. In the current experiment, it is worth noting that for both source videos more than half the subjects felt visual fatigue in some degree. Furthermore, a 15% of the observers felt headache or slight dizziness after seeing any of the 3D source sequences. Finally, some observers reported that other factors were also annoying, like the reduction of luminance while seeing 3D video wearing shutter glasses, or a very high degree of scene activity.

4.3. Preference between conventional and 3D video

After seeing the same sequence in monoscopic and stereoscopic formats, the observers were asked to indicate their preference between both. The obtained results are summarized in Table 3. More than half the observers preferred the stereoscopic version of the video

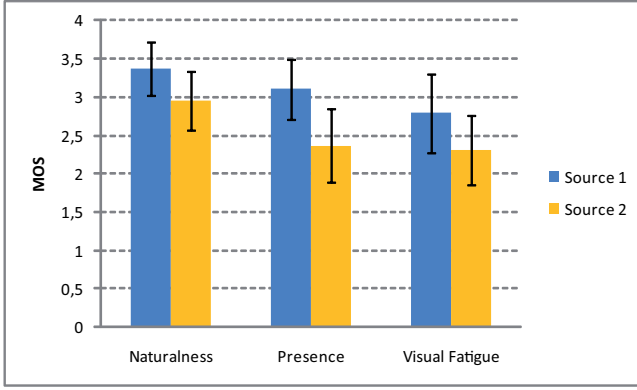


Fig. 4. Results for 3D performance.

source 1 to the monoscopic. In contrast, the majority of the observers preferred the monoscopic version of video source 2. Various factors influence the decisions of the observers, like the content of the sequences and a proper creation of the video. The results are in concordance to those showed in Figure 4 and mentioned in the previous Subsection, and indicate that the viewers will only switch to 3D technology when an added value with respect to viewing conventional video without causing discomfort.

Source	Preference (%)	
	3D	2D
1	52.63	47.37
2	21.05	78.95

Table 3. Results for preference between 2D and 3D presentations.

5. CONCLUSION

In this paper various aspects related to QoE for 3DTV were analyzed, from the results of subjective assessment tests. A single stimulus methodology was used, keeping as close as possible realistic home environment viewing conditions.

Firstly, the impact of transmission errors on the quality perceived by the end-users of an IPTV broadcasting system were analyzed. Typical errors related to packet-based networks were considered, like video and audio losses, bitrate and framerate drops and video freeze. Monoscopic and stereoscopic sequences were considered to compare the effects on each content. Side-by-Side multiplexing was used for 3D video, as it is currently used to broadcast 3D content to households. In both cases, video freezes combined with audio losses and those distortions affecting the natural flow of the video were the worst. However, it is worth noting the effects of video losses are more annoying in SbS 3D video, since they produce blocking artifacts which distort left and right views differently, and this cause binocular rivalry and important visual discomfort to the viewers. The analysis of the effects of the different distortions considered in the subjective experiment will be helpful for developing objective metrics, especially focused on monitoring video quality in packet-based networks used to deliver IPTV and 3DTV.

The performance of 3DTV was also evaluated rating the perceived naturalness, sense of presence and visual fatigue. Although these aspects are highly related to the video content of the sequences, it is worth noting that more than half the observers felt some kind of discomfort after viewing 3D content. In addition, the observers also indicated their preference between monoscopic and stereoscopic versions of the same content. The results showed that 3D video is only preferred when depth perception provides a substantial added value.

All these results show that the first generation of 3DTV broadcasting based on spatial multiplexing of left and right views is a good first approach to bring 3D video content to households. However, the research on more optimal coding, transmission and display techniques for 3D video is required to obtain better QoE and reduce visual discomfort.

6. ACKNOWLEDGMENTS

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under projects TEC2007-67764 (SmartVision) and TEC2010-20412 (Enhanced 3DTV), and by the Ministerio de Industria under project ITEA2-JEDI.

7. REFERENCES

- [1] P. Merkle, K. Muller, and T. Wiegand, "3D Video: Acquisition, Coding, and Display," *IEEE Trans. on Consumer Electronics*, vol. 56, no. 2, pp. 946–950, May 2010.
- [2] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *IEEE Trans. on Broadcasting*, vol. 54, no. 3, pp. 660–668, Sept. 2008.
- [3] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New Requirements of Subjective Video Quality Assessment Methodologies for 3DTV," *Int. Workshop on Video Processing and Quality Metrics*, Jan. 2010.
- [4] C. Hewage and M. Martini, "Reduced-Reference Quality Metric for 3D Depth Map Transmission," *Proc. 3DTV-CON*, June 2010.
- [5] T. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, and A. Reibman, "A Versatile Model for Packet Loss Visibility and its Application in Packet Prioritization," *IEEE Trans. on Image Processing*, vol. 19, no. 3, pp. 722–735, Mar. 2010.
- [6] P. Pérez, J. Macías, J.J. Ruiz, and N. García, "Effect of Packet Loss in Video Quality of Experience," *Bell Labs Technical Journal*, vol. 16, no. 1, pp. 91–104, June 2011.
- [7] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnström, R. Olsson, and P. Le Callet, "Subjective Quality Assessment of Error Concealment Strategies for 3DTV in the Presence of Asymmetric Transmission Errors," *Int. Packet Video Workshop*, Dec. 2010.
- [8] D. Strohmeier, S. Jumisko-Pyykkö, K. Kunze, G. Tech, D. Buğdayci, and M. OguzBici, "D4.3 Results of Quality Attributes of Coding, Transmission, and their Combinations," *Mobile3DTV Technical Report*, Feb. 2010, Available at <http://sp.cs.tut.fi/mobile3dvtv/>.
- [9] ITU, "Features of Three-Dimensional Television Video Systems for Broadcasting BT Series Broadcasting Service," *Rep. ITU-R BT.2160*, 2009.
- [10] ITU, "Methodology for the Subjective Assessment of the Quality of Television Pictures," *Rec. ITU-R BT. 500-11*, 2002.
- [11] VQEG, "Validation of Reduced-Reference and No-Reference Objective Models for Standard Definition Television, Phase I," 2009, Available at <http://www.vqeg.org>.
- [12] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites, "A Study of Visual Fatigue and Visual Comfort for 3D HDTV/HDTV Images," *Displays*, vol. 23, no. 4, pp. 191–201, Sept. 2002.
- [13] H. Coolican, *Research Methods and Statistics in Psychology*, Hodder & Stoughton, 2nd edition, 1994.